استناد به این مقاله: شریف، عاطفه (۱۳۹۱). بررسی تغییرات عملکرد دو موتور کاوش عمومی یاهو و گوگل از نظر پوشش کمّی- زمانی نمایه‌سازی و توجه به عناصر ابرداده‌ای در رتبه‌بندی صفحه‌های وب.
پژوهش‌نامه کتابداری و اطلاع‌رسانی،۲ (۱)، ۱۵۹-۱۷۸.

بررسی تغییرات عملکرد دو موتور کاوش عمومی یاهو و گوگل از نظر پوشش کمّی- زمانی نمایه‌سازی و توجه به عناصر ابرداده‌ای[1] در رتبه‌بندی صفحه‌های وب

عاطفه شریف[2]

دریافت: ۱۳۹۰/۷/۴ پذیرش: ۱۳۹۰/۱۰/۲۶

چکیده

هدف: *پژوهش حاضر با هدف ارزیابی عملکرد دو موتور کاوش عمومی گوگل و یاهو از نظر پوشش کمّی- زمانی نمایه‌سازی و توجه به پنج عنصر ابرداده‌ای -نشانه عنوان زبان نشانه‌گذاری فرامتن، ابرنشانه‌های کلیدواژه‌ها و توصیف زبان اچ تی ام ال، عناصر ابرداده‌ای عنوان، و موضوع از قالب ابرداده‌ای دابلین کور- که بر بازنمون موضوعی صفحه‌های وب متمرکزند، انجام شد.*

روش: *با استفاده از روش تجربی، نشریه‌ای الکترونیکی با ۹۳ صفحه، که ۸۴ صفحه آن مربوط به گروه‌های گواه و آزمون‌اند، به عنوان جامعه مورد مطالعه انتخاب شد. این صفحه‌ها در بازه زمانی نخست تحت دامنه فرعی http://metadata.irandoc.ac.ir و سپس در بازه زمانی دوم در دامنه فرعی http://metadata.um.ac.ir منتشر شد. روند نمایه‌سازی و معناداری تفاوت میان رتبه صفحه‌های دو گروه با استفاده از آزمون ناپارامتری یو من ویتنی محاسبه گردید. همچنین معناداری تغییر رتبه‌ها در دو بازه زمانی توسط آزمون آماری ویلکاکسون مورد بررسی قرار گرفت.*

یافته‌ها: *یافته‌ها نشان داد عنصر عنوان اچ تی ام ال در هر دو بازه زمانی اثربخش است. همچنین عنصر توصیف اچ تی ام ال در صورت منحصر به فرد بودن، در هر دو بازه زمانی عملکرد را بهبود بخشید. موتور کاوش گوگل نسبت به تکرار کلیدواژه در بخش سرآیند صفحه‌ها حساس است و آن را نوعی تقلب به شمار می‌آورد. همچنین اندک تفاوت‌هایی در عملکرد موتور کاوش گوگل نسبت به حضور عناصر ابرداده‌ای و تکرار آن‌ها، در طول زمان، وجود دارد.*

کلیدواژه‌ها: *تغییرات عملکرد، ابرداده، ابرنشانه‌های اچ تی ام ال، دابلین کور، بازیابی اطلاعات، رتبه‌بندی نتایج کاوش، موتورهای کاوش عمومی، دسترس پذیری صفحه‌های وب*

مقدمه

---

[1] این مقاله برگرفته ازطرحی پژوهشی است که به سفارش بنیاد ملی نخبگان و تحت نظارت پژوهشکده مطالعات فرهنگی و اجتماعی وزارت علوم، تحقیقات و فناوری در سال ۱۳۸۹ به انجام رسیده است.

۲ دانشجوی دکتری کتابداری و اطلاع رسانی دانشگاه فردوسی مشهد، atefehsharif@gmail.com

رشد سریع وب سایت‌ها، مراجعه فراوان کاربران به وب برای رفع نیازهای اطلاعاتی و بسیاری دلایل دیگر، متخصصان را بر آن داشته است تا به روش‌های گوناگون، دسترسی به اطلاعات انتشار یافته در وب را فراهم آورند. تلاش طراحان نظام‌های ذخیره و بازیابی در تولید و گسترش موتورهای کاوش نیز در این راستا است. هیچ یک از ابزارهای کاوش فعلی – از جمله موتورهای کاوش و راهنماها- قادر به پوشش تمامی مطالب منتشر شده در اینترنت نیستند(Henshaw and Valauskas, 2001, p.87) و شکاف میان تعداد صفحه‌های تولید شده و صفحه‌های نمایه‌سازی شده رو به افزایش است.

طراحان وب سایت‌ها با این هدف به طراحی صفحه‌ها می‌پردازند که این صفحه‌ها توسط کاربران بالقوه وب پیگیری شود و مورد مشاهده قرار گیرد. از آنجا که بیشتر کاربران از موتورهای کاوش بهره می‌گیرند، تنها زمانی می‌توان از بازیابی صفحه‌ها اطمینان حاصل کرد که اطلاعات نمایه‌ای صفحه‌ها در پایگاه اطلاعاتی موتورهای کاوش قرار گرفته باشد. مدتی طول می‌کشد تا موتور کاوش، صفحه‌ها را شناسایی نموده و اطلاعات لازم را در پایگاه اطلاعاتی خویش ذخیره سازد.

عملکرد موتورهای کاوش از جنبه‌های گوناگون قابل ارزیابی است. با این پیش فرض که موتور کاوش نمونه‌ای از یک نظام ذخیره و بازیابی است، شناسایی صفحه‌ها و نمایه‌سازی آنها، نخستین مرحله به شمار می‌رود. به این ترتیب مدت زمانی که صفحه‌ها مورد شناسایی قرار می‌گیرند و نمایه می‌شوند و همچنین عمق نمایه‌سازی - پوشش کمّی یا تعداد صفحه‌های یک سایت که مورد شناسایی قرار گرفته و نمایه می‌شود - حقایقی را درباره عملکرد موتورهای کاوش آشکار می‌سازد. جدا از مدت زمانی که نمایه‌سازی صفحه‌ها به طول می‌انجامد، چگونگی نمایه‌سازی محتوای صفحه‌ها نیز از اهمیت ویژه‌ای برخوردار است و می‌تواند یکی از معیارهای ارزیابی عملکرد موتورهای کاوش به شمار رود.

جزئیات چگونگی نمایه‌سازی صفحه‌های وب و عوامل اثرگذار بر آن به دلایل تجاری و جلوگیری از سوء استفاده‌های احتمالی مشخص نیست. همچنین توافقی بر سر مؤثرتر بودن عاملی نسبت به سایر عوامل وجود ندارد (Fishkin, 2005). علاوه بر این، به نظر می‌رسد که هر یک از موتورهای کاوش، الگوریتم رتبه‌بندی خاصی را دنبال می‌کنند (Henshaw and Valauskas, 2001, p. 92)؛ با وجود این، عوامل داخلی مانند ساختار ابرداده[1] و محتوای صفحه[2] و عوامل خارجی مانند وضعیت ارجاعات فرامتنی[3] می‌توانند بر رتبه‌بندی صفحه‌ها اثرگذار باشند (Zhang and Jastram, 2005).

---

[1] Metadata Structure
[2] Webpage Content
[3] Hyperlink Cited status

در سال ۱۳۸۶ پژوهشی با هدف تعیین اثربخشی عناصر ابرداده‌ای بر رتبه‌بندی صفحه‌های وب، توسط پژوهشگر انجام شد[۱]. در آن پژوهش ضمن ردگیری پوشش کمّی – زمانی نمایه‌سازی توسط موتورهای کاوش، میزان توجه این موتورها به عناصر ابرداده‌ای[۲] مورد بررسی قرار گرفت. ناپایداری، یکی از ویژگی‌های فضای مجازی است. در این وضعیت، انتظار می‌رود که موتور کاوش مانند هر نظام ذخیره و بازیابی دیگری در طول زمان به بهبود عملکرد خویش بپردازد و ضمن حفظ عملکردهای مثبت، در جهت کاهش کاستی‌ها تلاش کند.

مسئله آن است که آیا گذر زمان، تغییری در روند نمایه‌سازی و رتبه‌بندی صفحه‌های وب داشته است؟ و آیا تغییر رضایت‌بخشی در دو مؤلفه مورد توجه این پژوهش، یعنی مولفه کمّی بهبود پوشش کمّی – زمانی نمایه‌سازی و تغییر رویکرد در توجه به عناصر ابرداده‌ای، که به منزله مؤلفه‌ای کیفی مد نظر قرار گرفته، حاصل شده است؟

به این ترتیب پژوهش حاضر، با حفظ روش و صفحه‌های وب طراحی شده در پژوهش پیشین – شریف، ۱۳۸۶– عملکرد موتورهای کاوش را در بعد نمایه‌سازی و به لحاظ کمّی (کمّی- زمانی) و کیفی در دو بازه زمانی مورد بررسی قرار می‌دهد. در بعد کمّی- زمانی، سرعت و عمق نمایه‌سازی در موتورهای کاوش عمومی یاهو و گوگل مورد توجه است و در بعد کیفی، اثربخشی حضور عناصر ابرداده‌ای - به صورت منحصر به فرد و در صورت تکرار– به منزله معیار ربط فنی[۳] مورد توجه قرارگرفته‌اند. هر دو بعد کمّی- زمانی و کیفی با توجه به یافته‌های پژوهش شریف (۱۳۸۶) و با توجه به عملکرد همان موتور کاوش در مطالعه پیشین مورد تحلیل قرار می‌گیرد. عناصر ابرداده‌ای مورد بررسی در این پژوهش عبارتند از: ۱. نشانه عنوان[۴] از زبان اچ تی ام ال، ۲. ابرنشانه کلیدواژه‌ها[۵] از زبان اچ تی ام ال، ۳. ابرنشانه توصیف[۶] از زبان اچ تی ام ال، ۴. ابرداده عنوان[۷] از قالب ابرداده‌ای دابلین کور، و ۵. ابرداده موضوع[۸] از قالب ابرداده‌ای ابرداده‌ای دابلین کور.

---

۱ شریف، عاطفه (۱۳۸۶). بررسی میزان اثربخشی عناصر ابرداده‌ای بر رتبه‌بندی صفحات وب توسط موتورهای کاوش عمومی.کتابداری و اطلاع رسانی، ۲(۱۰)

2 Metadata elements
3 Technical Relevance
4 <title> </title>
5 <meta name=" keywords " content=" ">
6 <meta name=" description " content=" ">
7 <meta name="DC.Title" content=" ">
8 <meta name="DC.Subject" content=" ">

در پژوهش پیشین (شریف، ۱۳۸۶) سه موتور کاوش گوگل، یاهو، و ام اس ان[۱]، با توجه به معیارهای ویژه[۲]، انتخاب شدند. در این پژوهش، موتور کاوش ام اس ان (بینگ)، از چرخه ارزیابی عملکرد خارج شده است. علت این امر، پوشش اندک این موتور کاوش در نمایه‌سازی صفحه‌های طراحی شده در پژوهش پیشین است (تنها ۲۴ صفحه از ۹۳ صفحه طراحی شده را نمایه نموده است).

پرسش‌های پژوهش

هدف از انجام این پژوهش آگاهی نسبت به عملکرد نمایه‌سازی موتورهای کاوش از دو بعد کمّی– زمانی و کیفی در طول زمان است. به بیانی، سنجش تغییرات عملکردی موتورهای کاوش در نمایه‌سازی صفحه‌های وب در طول زمان ارزیابی می‌گردد. تغییرات حاصل شده در نمایه‌سازی، در رتبه صفحه‌های وب نمود می‌یابد. پاسخ به دو پرسش زیر معیاری برای ارزیابی موتورهای کاوش مورد مطالعه است:

۱. تغییرات عملکرد موتورهای کاوش از بعد پوشش کمّی- زمانی نمایه‌سازی در موتورهای کاوش مورد بررسی چگونه است؟

۲. تغییرات عملکرد موتورهای کاوش در اهمیت دادن به عناصر ابرداده‌ای برای رتبه‌بندی نتایج کاوش چگونه است؟

پرسش اول، بعد کمی و پرسش دوم، بعد کیفی را ارزیابی می‌نماید. برای دستیابی به پاسخ پرسش دوم دو پرسش جزئی به صورت زیر مطرح شد:

الف)عملکرد موتورهای کاوش در توجه به حضور کلیدواژه مورد کاوش در هر یک از عناصر مورد بررسی (نشانه عنوان، ابرنشانه‌های کلیدواژه‌ها و توصیف زبان نشانه‌گذاری فرامتن، ابرداده‌های عنوان و موضوع قالب ابرداده‌ای دابلین کور) چگونه است؟

ب)عملکرد موتورهای کاوش در توجه به تکرار کلیدواژه مورد کاوش در دو تا پنج عنصر یاد شده در بخش سرآیند[۳] صفحه و صرفنظر از نوع عنصر، چگونه است؟

پیشینه پژوهش

۱. در حال حاضر به نام Bing معروف است.

۲ در سال ۲۰۱۱ در سایت http://www.seoconsultants.com/search-engines همچنان سه موتور کاوش گوگل، یاهو و بینگ در صدر سیاهه‌ی موتورهای کاوش قرار دارند.

3 Header

با تمرکز بر ارزیابی موتور کاوش

مرور نوشتارها در داخل و خارج نشان می‌دهد که موتورهای کاوش از جنبه‌های متنوعی مورد پژوهش قرار گرفته‌اند. در مواردی رتبه‌بندی نتایج در انواع موتورهای کاوش از نظر میزان همپوشانی مورد توجه بوده است (Bar_Ilan, 2005)، گاه تفاوت رتبه‌بندی در طول زمان (Gotlieb and Eliopoulos, 2003)، تفاوت عملکرد موتورهای کاوش در روزآمدسازی نمایه‌های خود (Lewandowski, 2008؛ Notess, 2003؛ Lewandowski, Wahlig and Meyer-Bautor, 2006)، و گاه تفاوت عملکرد موتورهای کاوش عمومی با ابرموتورهای کاوش از جنبه تفاوت رتبه‌بندی و میزان همپوشانی مقایسه شده است (اسفندیاری مقدم، ۱۳۸۴؛ محمد اسماعیل و فیروزی، ۱۳۸۸). در بسیاری از پژوهش‌ها که مبنا را بر ارزیابی عملکرد موتورهای کاوش قرار داده‌اند، نتایج حاصل از کاوش به متخصصان موضوعی ارائه می‌شود. رتبه‌بندی متخصصان موضوعی با رتبه‌بندی حاصل از کاوش در موتور کاوش مقایسه و تفاوت میان این دو به عنوان معیار ارزیابی عملکرد موتور کاوش مورد استناد قرار می‌گیرد (Namjoshi, 2004؛ Vaughan, 2003). این دست از پژوهش‌ها در ارزیابی کیفیت نتایج کاوش در موتورهای کاوش عمومی و تخصصی نیز به انجام رسیده است (صراطی شیرازی، ۱۳۸۸).

با تمرکز بر ابرداده

پژوهش در حوزه ابرداده – در داخل و خارج - با دو رویکرد کلی انجام می‌پذیرد؛ ۱) رویکرد توصیفی؛ ۲) رویکرد تحلیلی.

در رویکرد توصیفی پژوهش‌هایی به عرصه کاربرد وارد شده، میزان استفاده از عناصر ابرداده‌ای را مورد سنجش کمی قرار داده‌اند (Drott, 2002؛ علیمحمدی، ۱۳۸۲؛ Craven, 2000, 2001a, 2001b؛ Qin and Wesley, 1998؛ Lawrence and Giles, 2000؛ Henshaw, 1999 ؛ شفیعی علویجه، غائبی و رضایی شریف آبادی، ۱۳۸۹). پیش فرض این دسته از پژوهش‌ها آن است که حضور عناصر ابرداده‌ای در بازیابی اطلاعات اثربخش است. اینگونه پژوهش‌ها با روش‌های کمی و از جنبه کاربردی انجام می‌پذیرد، جامعه کوچکی از صفحه‌های وب را مورد بررسی قرار می‌دهد و تا اندازه‌ای روشنگر میزان استفاده از عناصر ابرداده‌ای است.

در رویکرد تحلیلی نیز دو دسته پژوهش قابل بررسی است. دسته نخست پژوهش‌هایی هستند که با رویکردی تحلیلی به بررسی سطوح توصیف (Burnett, NG, and Park, 1999؛ تقی پور، ۱۳۸۴)، ارزیابی کیفی و چرایی عناصر توصیف در طرح‌های ابرداده‌ای می‌پردازند. این دسته از پژوهش‌ها با هدف تولید، توسعه و پالایش طرح‌های ابرداده‌ای انجام می‌پذیرد و نگاهی بر مبانی و اصول سازماندهی اطلاعات دارد.

دسته دوم پژوهش‌هایی هستند که با هدف سنجش میزان اثربخشی حضور عناصر ابرداده‌ای انجام می‌پذیرند. در این دسته از پژوهش‌ها، پیش فرض پژوهش‌هایی که تنها به سنجش کمی میزان استفاده از عناصر ابرداده‌ای می‌پردازند، مورد آزمون قرار می‌گیرد. در واقع تمرکز این گروه، تعیین اثربخشی در سطح نظام ذخیره و بازیابی اطلاعات به وسیله روش‌های تجربی است. برخی از این پژوهش‌ها یک نمونه را در دو مرحله قبل و بعد از افزودن عناصر ابرداده‌ای مورد بررسی قرار می‌دهند (Truner and Brackbill, 1998؛ Henshaw and Valauskas, 2001؛ Zhang and Dimitroff, 2005a, 2005b؛ صفری، ۱۳۸۲). در چنین وضعیتی عامل زمان خود مداخله‌گری است که می‌تواند بر نتایج پژوهش تأثیر گذار باشد. عواملی چون تغییر در صفحه‌های وب، پیوندهای فرامتنی به صفحه‌ها، و الگوریتم‌های رتبه‌بندی موتورهای کاوش (Bar-Ilan, 2005، ص ۱۵۱۸) از جمله تغییراتی است که احتمال وقوع آن در طول زمان وجود دارد. اما برخی دیگر از پژوهش‌های تجربی با دو نمونه و در یک مرحله انجام می‌پذیرند (Zhang and Dimitroff, 2004؛ Mohamed, 2006؛ شریف، ۱۳۸۶ ؛ طاهری، حریری و فتاحی، ۱۳۸۹) در این روش نقدی که بر پژوهش‌های دو مرحله‌ای وارد است تا حدودی ضعیف‌تر، و کنترل عوامل مداخله‌گر قوی‌تر است. زیرا دو گروه گواه و آزمون همزمان و تحت شرایط مشابهی در معرض نمایه‌سازی و سپس بازیابی قرار می‌گیرند.

انجام پژوهش‌های تحلیلی و توصیفی هر دو در جایگاه خویش حائز اهمیت هستند. اما ماندن در مرحله توصیف، به ارزیابی، تحلیل مشکلات، رفع موانع و بهبود سطح عملکرد نخواهد انجامید. بررسی‌های انجام شده در سطح نظام ذخیره و بازیابی می‌تواند در سطح کاربرد اثرگذارد؛ به این ترتیب که حصول اطمینان از اثربخشی عناصر ابرداده‌ای در سطح نظام، بر میزان و کیفیت استفاده در سطح کاربرد خواهد انجامید و بالعکس. جدول ۱ تصویری کلی از پژوهش‌های صورت گرفته در حوزه موتورهای کاوش و طرح‌های ابرداده‌ای می‌باشد.

جدول۱. دسته‌بندی پژوهش‌های مورد توجه در مرور پیشینه

	موضوع پژوهش	**نمونه‌ای از پژوهش‌ها**
موتورهای کاوش	مطالعه همپوشانی انواع موتورهای کاوش	بار-ایلان، ۲۰۰۵
	تفاوت رتبه‌بندی در طول زمان	گوتلیب و الیپولوس، ۲۰۰۳
	تفاوت در روزآمدسازی نمایه‌ها	لواندوسکی ،۲۰۰۸؛ نوتس، ۲۰۰۳؛ لواندوسکی، والیگ و مایر باوتر، ۲۰۰۶
	تفاوت رتبه‌بندی و همپوشانی موتورهای کاوش و ابرموتورها	اسفندیاری مقدم، ۱۳۸۴؛ محمد اسماعیل و فیروزی، ۱۳۸۸
	تفاوت رتبه‌بندی موتورهای کاوش و متخصصان موضوعی	نمجوشی، ۲۰۰۴؛ واگان، ۲۰۰۳
	مقایسه نتایج کاوش در موتورهای کاوش عمومی و	صراطی شیرازی، ۱۳۸۸

	تخصصی	
ابرداده	میزان استفاده از عناصر ابردادهای (رویکرد توصیفی)	دروت، ۲۰۰۲؛ علیمحمدی، ۱۳۸۲؛ کریون، ۲۰۰۰؛ ۲۰۰۱a؛ ۲۰۰۱b؛ کین و وسلی، ۱۹۹۸؛ لارنس و گیلس، ۲۰۰۰؛ هنشا، ۱۹۹۹؛ شفیعی علویجه، غائبی و رضایی شریف آبادی، ۱۳۸۹
	تحلیل سطوح توصیف و کیفیت عناصر توصیف در طرح‌های ابردادهای (رویکرد تحلیلی)	بارنت، ان جی، و پارک، ۱۹۹۹؛ تقی پور، ۱۳۸۴
	تعیین اثربخشی عناصر ابردادهای در سطح نظام ذخیره و بازیابی (رویکرد تحلیلی / با یک نمونه در دو مرحله)	ترنر و برک بیل، ۱۹۹۸؛ ژانگ و دیمیتروف، ۲۰۰۵a؛ ۲۰۰۵b؛ صفری، ۱۳۸۲؛ هنشا و والاسکاس، ۲۰۰۱
	تعیین اثربخشی عناصر ابردادهای در سطح نظام ذخیره و بازیابی (رویکرد تحلیلی / با دو نمونه در یک مرحله)	ژانگ و دیمیتروف، ۲۰۰۴؛ محمد، ۲۰۰۶؛ شریف، ۱۳۸۶ ؛ طاهری، حریری و فتاحی،۱۳۸۹

با توجه به جدول ۱ و مرور پیشینه‌ها مشخص می‌شود که بیشترین پژوهش‌ها در زمینه تعیین میزان استفاده از عناصر ابردادهای انجام شده است و تعداد اندکی از پژوهش‌ها به تعیین میزان اثربخشی عناصر ابردادهای پرداخته‌اند. روش‌های متفاوتی جهت تعیین میزان اثربخشی به خدمت گرفته شده و در برخی موارد نتایج متضادی از پژوهش‌ها حاصل گشته است.

تعداد اندک پژوهش‌ها در این زمینه (Alimohammadi, 2005 ؛Zhang and Dimitroff, 2004)، ضعف موجود در روش‌شناسی برخی پژوهش‌ها و عدم حصول نتایج ثابت، تأییدی بر ضرورت انجام پژوهش‌های بیشتر در این زمینه است. ضمن آن که مطابق با کاوش انجام گرفته از سوی پژوهشگر، پژوهشی که به ارزیابی چگونگی عملکرد موتورهای کاوش در توجه به عناصر ابردادهای و در طول زمان انجام گرفته باشد، بازیابی نشد.اگر چه مطالعاتی پیرامون عملکرد نمایه‌سازی کمّی صفحه‌های وب و توجه به روزآمدی صفحه‌ها در پایگاه اطلاعاتی موتورهای کاوش انجام شده است(Lewandowski, 2008؛ Notess, 2003؛ Lewandowski, Wahlig and Meyer-Bautor, 2006).

روش پژوهش

در این پژوهش، تغییرات عملکردی موتورهای کاوش، مورد پرسش است، لذا فرایند نمایه‌سازی به لحاظ کمّی و زمانی و همچنین توجه به عناصر ابردادهای، هر دو، مورد بررسی قرار می‌گیرد. از آن جایی که امکان دارد کمترین تغییر در صفحه‌های طراحی شده، فرایند نمایه‌سازی را دچار تغییر سازد، از صفحه‌های طراحی شده توسط شریف (۱۳۸۶) استفاده شد. جزئیات بیشتر در بخش مراحل اجرا آمده است.

پژوهش حاضر از نوع کاربردی است. با آگاهی از وجود عواملی که بر رتبه‌بندی صفحه‌ها

اثرگذارند، جهت اعمال متغیرهای مستقل، پنج عنصر ابرداده‌ای نشانه عنوان، ابرنشانه‌های کلیدواژه‌ها و توصیف زبان نشانه گذاری فرامتن، ابرداده‌های عنوان و موضوع قالب ابرداده‌ای دابلین کور، کنترل متغیرهای دخیل، و مشاهده تغییرات در متغیر وابسته، رتبه اسناد بازیابی شده از موتورهای کاوش در مورد هر یک از کلیدواژه‌ها، از روش تجربی یک مرحله‌ای استفاده شد.

مراحل اجرا و چگونگی گردآوری اطلاعات

مراحل اجرا و چگونگی گردآوردی داده‌های هر بخش از پژوهش به این شرح است:

۱. مجموعه‌ای از صفحه‌های وب که در پژوهش شریف (۱۳۸۶) مورد استفاده قرار گرفته بود، برای آزمون تغییرات عملکرد موتورهای کاوش مورد استفاده قرار گرفت. این صفحه‌ها که هر یک نشانگر مقاله‌ای[۱] است، در قالب وب سایت نشریه‌ای الکترونیکی به زبان انگلیسی، و توسط واژه‌پرداز ورد[۲] طراحی شد. به این مسئله توجه شده است که پیوند منطقی میان تمامی **۹۳** صفحه طراحی شده برقرار باشد. پیوندها از صفحه نخست وب سایت (سطح اول) آغاز می‌شود و با یک سطح به شماره نشریه و عنوان مقاله‌ها هر شماره از نشریه می‌رسد (سطح دوم)، با دنبال کردن پیوندی دیگر (انتخاب هر مقاله از سیاهه مقاله‌های یک شماره نشریه) متن کامل آن مقاله نمایش داده می‌شود (سطح سوم). پیوندهایی جهت حرکت در هر سه سطح در نظر گرفته شده است. همچنین برای افزایش امکان ردگیری و نمایه شدن صفحه‌ها توسط موتورهای کاوش، صفحه‌ای به عنوان نقشه سایت[۳] در نظر گرفته شده و پیوندهایی به تمامی صفحه‌های وب سایت در آن قرار گرفته است. این صفحه و همچنین صفحه اصلی وب سایت به طور هم زمان به موتورهای کاوش مورد بررسی معرفی گردید[۴].

۲. صفحه‌های طراحی شده، یک بار در بازه زمانی **۲۴** خرداد تا **۸** شهریور **۱۳۸۵**[۵] به مدت **۷۷** روز (**۱۱** هفته)[۶] و بار دیگر در بازه زمانی **۲۵** خرداد تا **۲۳** آبان **۱۳۸۹**[۱] به مدت **۱۵۴** روز (**۲۲** هفته) منتشر شد.

---

[۱] تعداد ۱۴ مقاله با موضوع ابرداده، وب معنایی و هستی‌شناسی‌ها از ۶ نشریه الکترونیکی دسترسی آزاد گزینش شد و با ساختاری یکنواخت در قالب وب‌سایت نشریه قرار گرفت.

[2] Word

[3] Sitemap

[۴] Submit امکان معرفی وب‌سایت‌هایی که به تازگی انتشار یافته‌اند، در موتورهای کاوش وجود دارد. پژوهشگر جهت سرعت بخشیدن به فرایند شناسایی صفحات طراحی شده در موتورهای کاوش گوگل و یاهو، از این امکان موتورها استفاده کرده است.

[5] June 14, 2006 till Aug 30, 2006

[۶] تحلیل یافته‌های حاصل از دور نخست انتشار صفحات با تمرکز بر اثربخشی حضور عناصرابرداده‌ای بر رتبه‌بندی صفحات در شریف (۱۳۸۶)

انتشار صفحه‌های وب در بازه زمانی نخست در دامنه فرعی ابرداده[2] از سایت پژوهشگاه اطلاعات و مدارک علمی ایران[3] انجام پذیرفت و انتشار همان صفحه‌ها، در بازه زمانی دوم در دامنه فرعی ابرداده از سایت دانشگاه فردوسی مشهد[4] صورت گرفت.

۳. یکی از معیارهای موتورهای کاوش در سرعت بخشیدن به فرایند نمایه‌سازی صفحه‌ها، اهمیت دامنه‌ای است که صفحه‌ها در آن به انتشار در می‌آیند. از مهمترین ملاک‌های اهمیت دامنه، تعداد پیوندهای دریافتی آن دامنه می‌باشد[5]. برای کنترل این عامل مداخله‌گر، کاوشی در موتور کاوش گوگل به انجام رسید و مشخص گردید که میزان پیوندهای دریافتی وب سایت پژوهشگاه اطلاعات و مدارک علمی ایران بیش از وب سایت دانشگاه فردوسی مشهد است (۴۰۰ پیوند در مقابل ۲۷۰ پیوند). به همین دلیل و برای مهار این عامل مداخله‌گر، در بازه دوم انتشار صفحه‌ها در دامنه فرعی دانشگاه فردوسی، مدت زمانی طولانی‌تر – دو برابر بازه زمانی نخست – برای نمایه‌سازی در نظر گرفته شد.

۴. روند نمایه‌سازی صفحه‌ها در دو موتور کاوش ذکر شده، مورد مشاهده قرار گرفت و تغییرات به صورت روزانه ثبت شد[6]. از آن جایی که در موتور کاوش گوگل تعداد نتایج اعلام شده با آن چه که در واقعیت بازیابی شده است متفاوت است، تا انتهایی‌ترین صفحه نتایج، بررسی و تعداد واقعی نتایج، ثبت شد. تا این مرحله داده‌های مربوط به بخش سنجش عملکرد موتور کاوش در بعد کمّی و زمانی گردآوری شد[7].

گزارش شده است.

[1] June 13, 2010 till Nov 15, 2010

[2] Subdomain

[3] http://metadata.irandoc.ac.ir

[4] http://metadata.um.ac.ir

[5] منظور از پیوند دریافتی، تعداد پیوندهایی است که از سایر وب‌سایت‌ها به وب‌سایت مورد نظر برقرار شده است. تعداد پیوند دریافتی می‌تواند نشانی از اهمیت یک وب‌سایت در جامعه مخاطبان آن باشد. به طور مثال اگر به دست آوردن تعداد پیوندهای دریافتی به وب سایت دانشگاه فردوسی مشهد در موتور کاوش گوگل مورد نیاز باشد با کاوش فرمول link:www.um.ac.ir این تعداد مشخص می شود. در سایر موتورهای کاوش نیز وضعیت مشابهی حاکم است.

[6] جهت گردآوری داده‌ها در بازه زمانی نخست، روزی سه کاوش در موتور کاوش گوگل، یاهو و ام اس ان با فرمول site:http://metadata.irandoc.ac.ir انجام شد. در بازه زمانی دوم نیز روزانه دو کاوش در دو موتور کاوش گوگل و یاهو با فرمول site:http://metadata.um.ac.ir انجام پذیرفت و همچون گذشته، تعداد نتایج بازیابی شده ثبت گردید.

[7] فرایند گردآوری داده‌های مربوط به فرایند نمایه‌سازی در هر دو بازه زمانی، از یک IP واحد انجام پذیرفت؛ به طوری که در مرحله نخست، کاوش‌ها در دانشگاه تهران و در مرحله دوم، کاوش‌های مور د نظر در دانشگاه فردوسی مشهد انجام پذیرفته است. همچنین به منظور کاهش خطا، تلاش شد تا گردآوری داده‌های مرتبط با اثربخشی عناصر ابرداده‌ای که مبتنی بر رتبه‌بندی صفحات بوده است، به صورت فشرده

۵. پس از آن که تعداد صفحه‌های نمایه شده در موتورهای کاوش به حد قابل قبولی رسید برای تعیین عملکرد موتورهای کاوش در توجه آن‌ها به عناصر ابرداده‌ای کاوش کلیدواژه‌ای در کادر محاوره‌ای ساده دو موتور کاوش گوگل و یاهو انجام پذیرفت. از آنجایی که دو گروه صفحه‌های گواه و آزمون، در وب سایت طراحی شده، گنجانده شده است؛ و نیازی به تعیین رتبه صفحه‌ها در میان سایر صفحه‌های موجود در وب نیست، کاوش کلیدواژه‌ای در دامنه وب سایت طراحی شده محدود شد. فرمول کاوش در موتور کاوش گوگل و یاهو به ترتیب چنین است:

site:http://metadata.irandoc.ac.ir “Keywords”

site:metadata.irandoc.ac.ir “Keywords”یا Domain:metadata.irandoc.ac.ir “Keywords”

تعداد صفحه‌های نمایه شده در موتور کاوش یاهو در انتهای بازه زمانی دوم با افتی ناگهانی مواجه شد. به این ترتیب عملکرد دو موتور کاوش گوگل و یاهو در مورد پوشش کمّی و زمانی نمایه‌سازی مورد تحلیل قرار گرفت و معیار کیفی ارزیابی، یعنی توجه به عناصر ابرداده‌ای، تنها در مورد موتور کاوش گوگل بررسی شد. به طوری که تفاوت عملکرد این موتور کاوش در مقایسه با داده‌های جمع‌آوری شده در بازه زمانی نخست از موتور کاوش گوگل، مورد آزمون قرار گرفت.

شیوه تجزیه و تحلیل داده‌ها

برای پاسخگویی به پرسش نخست که تغییرات عملکردی موتورهای کاوش را درباره پوشش زمانی و کمّی نمایه‌سازی در نظر دارد، داده‌های حاصل شده تا سومین گام گردآوری داده‌ها مورد تحلیل قرار گرفت. در مورد پرسش دوم برای تعیین اثربخشی عناصر ابرداده‌ای بر رتبه‌بندی موتورهای کاوش، رتبه حاصل از هر یک از کاوش‌های انجام شده در موتورهای کاوش، به تفکیک گروه گواه و آزمون با توجه به پرسش‌های پژوهش در SPSS وارد شد و مورد تجزیه و تحلیل قرار گرفت.

با توجه به این که نتایج حاصل از کاوش در موتورهای کاوش از نوع داده‌های رتبه‌ای است (Truner and Brackbill, 1998؛ Safari, 2005)، میزان معنی‌داری تفاوت میان گروه گواه و آزمون – دو گروه مستقل – در بازیابی صفحه‌های وب، با استفاده از آزمون ناپارامتری یو من ویتنی تعیین گردید. این آزمون در هر یک از دو بازه زمانی تکرار شد. برای تکمیل تحلیل‌ها، از آن جایی که تغییرات احتمالی عملکرد موتورهای کاوش مورد توجه است، آزمون ویلکاکسون در مورد رتبه‌های به دست آمده در هر یک از

و در یک روز انجام شود.

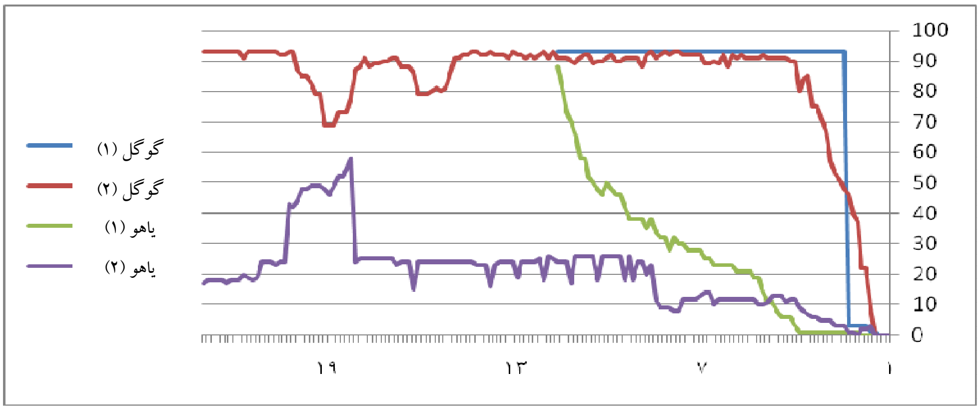
بازه‌های زمانی برای دو گروه صفحه‌های گواه و آزمون اجرا شد. به این ترتیب که معنی‌داری تفاوت میان عملکرد موتور کاوش در دو زمان T_1 و T_2 برای صفحه‌های آزمون و گواه مورد محاسبه قرار گرفت.

تحلیل و نتیجه‌گیری

۱. تغییرات عملکردی موتورهای کاوش از بعد پوشش کمّی- زمانی نمایه‌سازی در موتورهای کاوش مورد بررسی چگونه است؟

برای پاسخگویی به این پرسش، روند نمایه‌سازی در موتورهای کاوش در دو بازه زمانی مورد پیگیری قرار گرفت و به صورت روزانه ثبت گردید. عمل گردآوری داده‌های این بخش در نخستین بازه زمانی مدت ۱۱ هفته و در بازه زمانی دوم ۲۲ هفته به طول انجامید. در هر دو بازه زمانی، تنها موتور کاوش گوگل، تمامی صفحه‌های نشریه الکترونیکی طراحی شده را نمایه‌سازی کرد به طوری که در بازه زمانی نخست این فرایند در طی ۱۱ روز و در بازه زمانی دوم در طی ۷ هفته تکمیل شد.

فرایند نمایه‌سازی در موتور کاوش یاهو، روندی کندتر داشت؛ به طوری که در بازه زمانی نخست و در آخرین هفته، تعداد صفحه‌های نمایه‌سازی شده به ۸۸ عدد از ۹۳ صفحه رسید و در بازه زمانی دوم اگر چه این تعداد در هفته هجدهم به ۵۸ صفحه رسید، اما با طی روندی نزولی به تعداد ۱۸ عدد در پایان هفته بیست و دوم رسید.

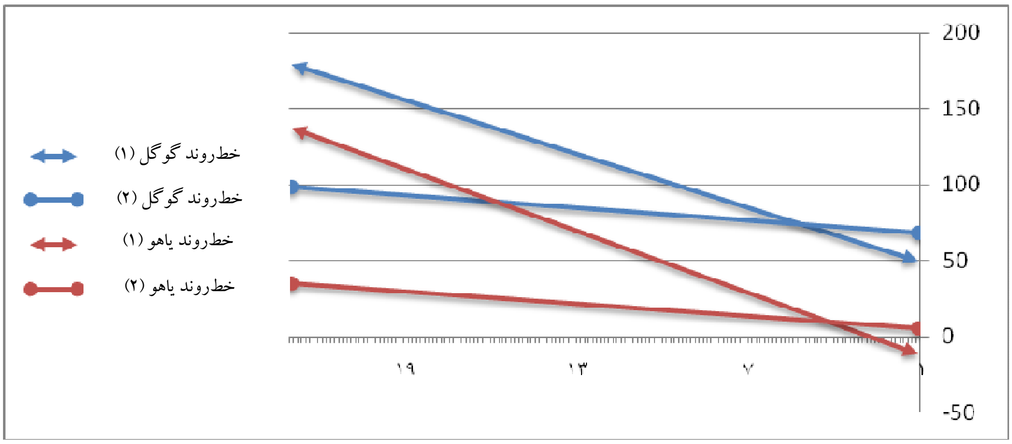


نمودار ۱. وضعیت کمّی و زمانی نمایه‌سازی صفحه‌ها در دو بازه زمانی در هر دو موتور کاوش گوگل، یاهو

در نمودار ۱، روند نمایه‌سازی دو موتور کاوش مورد بررسی در هر دو بازه زمانی آمده است. افت و خیزها در هر دو موتور کاوش در بازه زمانی دوم بسیار بیشتر از افت و خیزهای ملایم در بازه زمانی اول

است. ضمن آن که در موتور کاوش گوگل و در بازه زمانی نخست هیچ‌گونه افت و خیزی مشاهده نمی‌شود.

در نمودار۲، خط روندهای هر یک از موتورهای کاوش در دو بازه زمانی انتشار آمده است برای سهولت در تشخیص، خط روندها در مورد هر یک از موتورهای کاوش با رنگی یکسان نشان داده است. با دقت در نمودار ۲، روند نمایه‌سازی هر یک از موتورهای کاوش در مقایسه با موتور کاوش دیگر و همچنین در مقایسه با همان موتور کاوش در دو بازه زمانی متفاوت قابل تحلیل است.



نمودار ۲. خطوط روند نمایه‌سازی صفحه‌ها در دو بازه زمانی در هر دو موتور کاوش گوگل، یاهو

خط روند موتور کاوش یاهو در مرحله نخست بیشترین شیب را نشان می‌دهد و این نشان از فراز و فرودهای قابل ملاحظه‌ای است که پیشتر نیز به آن اشاره شد. همچنین مشاهده می‌شود که خط روند دو موتور کاوش گوگل و یاهو در هر یک از بازه‌های زمانی با فاصله‌هایی تقریبا مشابه از هم و به صورت موازی (البته نه کاملا) قرار گرفته‌اند (دو خط پیکان دار با هم و دو خط دایره دار با هم مقایسه شود). با وجود تفاوت‌هایی که در بعد پوشش کمّی و زمانی نمایه‌سازی مورد اشاره قرار گرفت، خط روندهای موازی، نشانگر ثبات نسبی عملکرد هر یک از موتورهای کاوش در فرایند نمایه‌سازی صفحه‌ها است.

۲. تغییرات عمکردی موتورهای کاوش در اهمیت دادن به عناصر ابردادهای در رتبه‌بندی نتایج کاوش چگونه است؟

۲-۱. اثربخشی حضور هر یک از عناصر ابردادهای در موتور کاوش گوگل

برای درک بهتر و امکان مقایسه و تحلیل نتایج، جدول۲ که نمایی کلی از مقدار پی محاسبه شده در

آزمون یو من ویتنی است، ارائه شد. موارد معنی‌دار هر عنصر، با قلم درشت و ایتالیک مشخص شده است. تفاوت میانگین رتبه‌های کسب شده در هر گروه نشانگر آن است که رتبه به نفع کدامین گروه تغییر وضعیت داده است؛ لذا در هر مورد از عناصر، به تفکیک موتور کاوش، گروه صفحه‌های که بهبود رتبه داشته – آزمون یا گواه- با اصطلاح "به نفع" مشخص شده است. از آن جایی که تعداد صفحه‌های نمایه شده در موتور کاوش یاهو کمتر از حد انتظار بود، لذا در تحلیل‌ها، تنها بر موتور کاوش گوگل تمرکز شده است.

جدول ۲. خلاصه نتایج آزمون یو من ویتنی در چگونگی عملکرد موتور کاوش گوگل در اهمیت دادن به عناصر ابرداده‌ای

عنصر مورد آزمون	**نوع کلیدواژه**	**بازه زمانی نخست**		**بازه زمانی دوم**	
		P-value	**به نفع**	P-value	**به نفع**
عنوان اچ تی ام ال	کلیدواژه منحصر به فرد	--	--	--	--
	کلیدواژه تکرار شونده	***۰/۰۰۰***	آزمون	***۰/۰۰۰***	آزمون
کلیدواژه‌های اچ تی ام ال	کلیدواژه منحصر به فرد	۰/۴۶۲	آزمون	۰/۶۸۵	آزمون
	کلیدواژه تکرار شونده	۰/۴۷۲	آزمون	***۰/۰۰۶***	گواه
توصیف اچ تی ام ال	کلیدواژه منحصر به فرد	***۰/۰۱۵***	آزمون	***۰/۰۱۴***	آزمون
	کلیدواژه تکرار شونده	***۰/۰۱۶***	گواه	***۰/۰۲۵***	آزمون
عنوان دابلین کور	کلیدواژه منحصر به فرد	--	--	--	--
	کلیدواژه تکرار شونده	۰/۸۲۶	آزمون	۰/۵۳۷	گواه
موضوع دابلین کور	کلیدواژه منحصر به فرد	۰/۱۲۱	آزمون	۰/۰۷۶	آزمون
	کلیدواژه تکرار شونده	***۰/۰۳۸***	گواه	۰/۱۱۷	*گواه*

با توجه به یافته‌های جدول۲ می‌توان گفت:

- در هر دو بازه زمانی عنصر عنوان اچ تی ام ال برای موتور کاوش گوگل اهمیت داشته است.
- موتور کاوش گوگل نسبت به حضور عنصر ابرداده‌ای کلیدواژه‌های زبان اچ تی ام ال، در صورت منحصر به فرد بودن، بهبود عملکردی نداشته است. این روند در طی هر دو بازه زمانی ثابت بوده است. در مورد کلیدواژه‌های تکراری و در بازه زمانی دوم، حضور کلیدواژه‌های تکرار شونده نه تنها به بهبود عملکرد نیانجامیده است، بلکه باعث افت رتبه صفحه‌ها بازیابی شده نیز شده است.
- عنصر توصیف اچ تی ام ال در صورت منحصر به فرد بودن، در هر دو بازه زمانی به بهبود عملکرد انجامیده است. در مورد کلیدواژه‌های تکرار شونده، اگر چه در بازه زمانی نخست، حضور این کلیدواژه‌ها باعث افت رتبه‌ها شده است در بازه زمانی دوم به نفع گروه آزمون به بهبود رتبه انجامیده است.

- عنصر ابرداده‌ای دابلین کور مورد توجه موتور کاوش گوگل نبود.
- کلیدواژه‌های منحصر به فرد در مورد عنصر ابرداده‌ای موضوع دابلین کور، تفاوت معناداری را در رتبه صفحه‌های گروه گواه و آزمون ایجاد نکرد. اما چنین به نظر می‌رسد که کلیدواژه‌های تکرار شونده در هر دو بازه زمانی باعث افت رتبه‌ها شده است. این تفاوت، در بازه زمانی نخست، معنادار است و در بازه زمانی دوم به سطح معناداری نرسیده است.

یافته‌های حاصل از بررسی اثربخشی دو ابرنشانه کلیدواژه‌ها و توصیف در این پژوهش، با یافته‌های ترنر و برک بیل (۱۹۹۸) مبنی بر اثربخشی ابرنشانه کلیدواژه‌ها در مقایسه با ابرنشانه توصیف متفاوت است. یافته‌های پژوهش حاضر نشان داد که ابرنشانه توصیف در مقایسه با ابرنشانه کلیدواژه‌ها اثربخشی بیشتری در بهبود رتبه صفحه‌ها در دو بازه زمانی داشته است. در این اثربخشی، کلیدواژه‌های تکرار شونده نیز باعث بهبود رتبه‌ها شده است.

همچنین یافته‌ها عدم اثربخشی عناصر ابرداده‌ای دابلین کور در رتبه‌بندی نتایج کاوش را، آن گونه که در پژوهش صفری (۲۰۰۵) و شریف (۱۳۸۶) گزارش شده است، تأیید می‌کند. این در حالی است که ژانگ و دیمیتروف (۲۰۰۴) اثربخشی عنصر موضوع دابلین کور را گزارش کرده‌اند.

تغییرات عملکردی موتور کاوش گوگل در طول زمان از نظر توجه به عناصر ابرداده‌ای مورد بررسی

با توجه به جدول۲، در مورد تغییرات عملکرد موتور کاوش گوگل در طول زمان می‌توان گفت که تنها در دو مورد تفاوت عملکرد مشاهده می‌شود. نخست در مورد کلیدواژه‌های تکرار شونده ابرنشانه کلیدواژه‌های زبان اچ تی ام ال، به نحوی که در بازه زمانی دوم، تفاوتی معنادار به نفع گروه گواه را نشان می‌دهد و دوم در کلیدواژه‌های تکرار شونده توصیف زبان اچ تی ام که نشانگر تغییر جهتی به سمت بهبود رتبه در جهت گروه آزمون است. در بقیه موارد، آن چنان که در جدول۲ آمده است، نوعی ثبات عملکرد حاکم می‌باشد. برای روشن‌تر شدن تغییرات عملکرد موتور کاوش گوگل در طول زمان، جدول۳ که گزارش آزمون ویلکاکسون است، ارائه می‌شود.

جدول ۳. نتایج آزمون ویلکاکسون در مورد تغییرات عملکرد موتور کاوش گوگل در توجه به عناصر

عنصر مورد بررسی	P-value	
	صفحه‌های گروه گواه	صفحه‌های گروه آزمون
عنوان اچ تی ام ال	۰/۴۴۰	۰/۱۵۸
کلیدواژه‌های اچ تی ام ال	۰/۰۶۰	۰/۲۱۶

توصیف اچ تی ام ال	۰/۵۸۶	***۰/۰۰۱***
عنوان دابلین کور	۰/۲۲۴	۰/۸۶۱
موضوع دابلین کور	***۰/۰۰۱***	***۰/۰۰۲***

همانطور که در جدول۳ مشخص است نخستین تفاوت معناداری که در طول زمان گزارش شده است مربوط به عنصر توصیف اچ تی ام ال است. بیان شد که در بازه زمانی نخست وجود عناصر ابردادهای در قالب کلیدواژه‌های تکرار شونده در توصیف اچ تی ام ال باعث افت معنادار رتبه‌ها و در بازه زمانی دوم سبب بهبود معنادار رتبه‌ها شده است (جدول۲). تفاوت معنادار حاصل از آزمون ویلکاکسون به این تغییر رویه در موتور کاوش گوگل مربوط است.

دومین تفاوت در مورد عنصر موضوع دابلین کور گزارش شده است. در این مورد نیز با توجه به جدول۲ می‌توان چنین استدلال کرد که اگر چه موتور کاوش گوگل، در مورد کلیدواژه‌های تکرار شونده به طور معناداری باعث افت رتبه‌ها شده است، در بازه زمانی دوم این تفاوت معنادار نیست. این تغییر رویه سبب معنادار شدن تفاوت عملکرد موتور کاوش گوگل در طول زمان شده است.

۲-۲. اثربخشی تکرار کلیدواژه در بخش سرآیند صفحه وب در موتور کاوش گوگل

در این پرسش تعداد تکرار کلیدواژه در بخش سرآیند صفحه‌های وب و اثرگذاری آن بر رتبه صفحه‌ها مورد بررسی قرار گرفته است.

با توجه به جدول۴ مشخص است که در هر دو بازه زمانی، هنگامی که تفاوت معناداری در رتبه‌ها مشاهده شده است، این تفاوت رتبه به نفع گروه گواه است، نه گروه آزمون. به این معنی که موتور کاوش گوگل نسبت به تکرار کلیدواژه در بخش سرآیند صفحه‌ها حساس است و آن را نوعی تقلب به شمار می‌آورد. نتیجه آزمون یو من ویتنی به ویژه در تکرار پنج باره کلیدواژه در بخش سرآیند، گواهی بر این مطلب است.

جدول۴. اثربخشی دو تا پنج بار تکرار کلیدواژه در بخش سرآیند صفحه وب بر رتبه‌بندی صفحه‌های وب در موتور کاوش گوگل

تعداد تکرار مورد بررسی	**بازه زمانی نخست**		**بازه زمانی دوم**	
	P-value	به نفع	**P-value**	به نفع
دوبار تکرار	۰/۵۸۰	آزمون	۰/۰۰۸	گواه

سه بار تکرار	۰/۰۰۸	گواه	۰/۶۶۳	آزمون
چهار بار تکرار	۰/۰۶۲	گواه	۰/۵۶۲	گواه
پنج بار تکرار	۰/۰۰۰	گواه	۰/۰۱۱	*گواه*

تغییرات عملکردی موتور کاوش گوگل در طول زمان از نظر توجه به تکرار عناصر ابرداده‌ای در بخش سرآیند صفحه‌های وب

در مورد تکرار کلیدواژه‌های تکرار شونده در بخش سرآیند صفحه‌های وب، با توجه به جدول۵، به نظر می‌رسد که تنها در تکرار پنج‌باره کلیدواژه، تفاوتی معنادار به نفع گروه گواه مشاهده می‌شود. در مورد دوبار تکرار کلیدواژه نیز تفاوتی میان دو بازه زمانی قابل شناسایی است به طوری که در بازه زمانی دوم، تفاوتی معنادار به نفع گروه گواه گزارش شده است.

اما آزمون ویلکاکسون در جدول۵ نشان می‌دهد که دو مورد تفاوت معنادار در عملکرد موتور کاوش گوگل در اهمیت دادن به تکرار کلیدواژه‌ها در بخش سرآیند مشاهده می‌شود. نخستین تفاوت معنادار مربوط به دو بار تکرار کلیدواژه‌ها در بخش سرآیند است. همانگونه که در جدول ۴ نیز مشاهده شد یک تغییر در موتور کاوش گوگل قابل شناسایی است. به طوری که در دومین بازه زمانی تفاوتی معنادار به نفع گروه گواه نشان داده است در حالی که در بازه زمانی نخست اگر چه مقدار پی محاسبه شده به سطح معناداری نرسیده است اما مجموع رتبه‌ها به نفع گروه آزمون بوده است. همین اتفاق به طور برعکس در مورد سه بار تکرار نیز مشاهده می‌شود.

جدول ۵. تغییر عملکرد موتور کاوش گوگل در توجه به تکرار دو تا پنج باره کلیدواژه در بخش سرآیند صفحه

تعداد تکرار مورد بررسی	P-value	
	صفحه‌های گروه گواه	**صفحه‌های گروه آزمون**
دوبار تکرار	۰/۱۵۸	۰/۰۰۱
سه بار تکرار	۰/۱۵۸	۰/۰۰۲
چهار بار تکرار	۰/۱۵۸	۰/۱۸۲
پنج بار تکرار	۰/۱۵۸	۰/۰۸۰

سخن پایانی

کوتاه سخن آن که عملکرد موتورهای کاوش از جنبه‌های مختلف قابل ارزیابی است. در این

پژوهش دو مولفه کمّی و کیفی در عملکرد نمایه‌سازی موتورهای کاوش یاهو و گوگل مورد توجه است. در بعد کمّی- زمانی، سرعت و عمق نمایه‌سازی در موتورهای کاوش عمومی مورد بررسی قرار گرفت و در بعد کیفی عناصر ابردادهای - که از جمله عوامل بالقوه مؤثر بر رتبه‌بندی نتایج کاوش هستند – به منزله معیار ربط فنی. در تحلیل یافته‌ها مشخص شد که به جز برخی موارد مختصر ذکر شده، موتورهای کاوش مورد بررسی تفاوت چندانی در عملکرد خویش نداشته‌اند.

این وضعیت سبب طرح پرسش‌های دیگری می‌شود. از جمله: در صورت اثربخش نبودن به کارگیری عناصر ابرداده‌ای، آیا لزومی به توسعه و استفاده از آنها در محیط وب هست؟ چه عوامل دیگری در کسب رتبه‌های بهتر در نتایج کاوش موتورها مؤثر است؟ تغییرات عملکردی موتورهای کاوش در چه سمت و سویی است؟ پاسخ اینگونه پرسش‌ها نیازمند طراحی پژوهش‌های دیگری است. به نظر می‌رسد که نیاز به ارزیابی همه جانبه این ابزارهای کاوش در جنبه‌های مختلف و با رویکردهای متفاوت همچنان وجود دارد.

کتابنامه

اسفندیاری مقدم، علیرضا (۱۳۸۴). بررسی نتایج جستجو در ابرموتورهای کاوش و موتورهای تحت پوشش آنها از جنبه همپوشانی و رتبه‌بندی . پایان‌نامه کارشناسی ارشد کتابداری و اطلاع‌رسانی، دانشکده روان شناسی و علوم تربیتی، دانشگاه فردوسی مشهد.

تقی پور، مژگان (۱۳۸۴). بررسی تطبیقی عناصر ابرداده‌های عمومی صفحه‌های وب. پایان‌نامه کارشناسی ارشد کتابداری و اطلاع‌رسانی، دانشکده روان شناسی و علوم تربیتی، دانشگاه تهران.

شریف، عاطفه (۱۳۸۶). بررسی میزان اثربخشی عناصر ابرداده‌ای بر رتبه‌بندی صفحه‌های وب توسط موتورهای کاوش عمومی. *کتابداری و اطلاع رسانی*، ۱۰ (۲)، ۲۴۱-۲۵۸.

شفیعی علویجه، سیما، غائبی، امیر، رضایی شریف آبادی، سعید (۱۳۸۹) بررسی عناصر ابرداده‌ای موجود در صفحه‌های وب حاصل از جستجو در موتورهای جستجوی عمومی. *علوم و فناوری اطلاعات*، شماره ۵۹، ۷۱-۸۹. دسترس پذیر در

http://www.magiran.com/view.asp?Type=pdf&ID=692690

صراطی شیرازی، منصوره (۱۳۸۸). بررسی مقایسه ای میزان دقت موتورهای کاوش عمومی و تخصصی پزشکی در بازیابی مدارک مربوط به بیماری های کودکان. *فصلنامه کتاب*، ۷۷

صفری، مهدی (۱۳۸۲). سنجش میزان اثربخشی عناصر ابرداده‌ای دابلین کور در بازیابی صفحه‌های وب: مطالعه صفحه‌های وب Iranian International Journal of Science . پایان‌نامه کارشناسی ارشد کتابداری و اطلاع‌رسانی، دانشکده روان شناسی و علوم تربیتی، دانشگاه تهران.

طاهری، مهدی، حریری، نجلا، فتاحی، رحمت الله (۱۳۸۸). بررسی تطبیقی کیفیت نمایه‌سازی و رتبه‌بندی اشیاء محتوایی حاوی عناصر فراداده‌ای هسته دوبلین و مارک ۲۱ توسط موتورهای کاوش عمومی. *کتابداری و اطلاع رسانی*. ۱۲(۴). دسترس پذیر در

http://www.aqlibrary.org/index.php?module=TWArticles&file=index&func=view_pubarticles&did=739&pid=10

علیمحمدی، داریوش (۱۳۸۲). ارزیابی ابرنشانه‌های وب سایت‌های ایران. پایان‌نامه کارشناسی ارشد کتابداری و اطلاع‌رسانی، دانشکده روان‌شناسی و علوم تربیتی، دانشگاه تهران.

محمد اسماعیل، صدیقه؛ فیروزی، صغری (۱۳۸۸) مقایسه میزان همپوشانی نتایج بازیابی شده در موتورهای کاوش و ابر موتورهای کاوش در بازیابی اطلاعات نانوتکنولوژی. *ماهنامه اطلاع‌یابی و اطلاع‌رسانی*، ۲۳

Alimohammadi, D. (2005). Meta-tags: still a matter of opinion. The Electronic Library, 23(6), 625-631.

Bar_Ilan, J. (2005). Comparing Rankings of Search Results on the Web. Information Processing and Management, 41. Retrieved Oct, 12, 2010, From Elsevier Database

Burnett, K.; Ng, K.B.; and Park, S. (1999). A Comparison of the Two Traditions of Metadata Development. Journal of the American Society for Information Science, 50 (13), 1209-1217.

Craven, T.C. (2000). Features of Description meta tags in Public home pages . Journal of Information Science, 26(5), 303-313.

Craven, T.C. (2001a). Changes In meta tags description over time. First Monday, 6(10). Retrieved Oct, 12, 2010, From http://firstmonday.org/issues/issue6_10/craven/index.html

Craven, T.C. (2001b). Description meta tags in Locally linked web pages. ASLIB Proceedings, 53(6).

Drott, M.C. (2002). Indexing aids at Corporate websites: the use of robot.txt and meta tags. Information Processing and Management, 38(2), 209-219.

Fishkin, R. (2005). Search Engine Ranking Factors Retrieved Des, 17, 2010, From http://www.seomoz.org/articles/search-ranking-factors.php#1

Gotlieb, C. ; Eliopoulos, D. (2003). Evaluating Web Search Results Ranking . Online, 27(2). Retrieved Oct, 11, 2009, From Ebsco Computer Science Database.

Henshaw, R. and Valauskas, E.J. (2001). Metadata as a catalyst: experiments with metadata and search engines in the internet journal, First Monday . Libri, 51(2), 86-101pp. Retrieved Oct, 23, 2009, From www.librijournal.org/pdf/1999-3pp125-131.pdf

Lawrence, S. and Giles, C.L. (2000). Accessibility of Information on the Web. Intelligence, 11(1), 32-39.

Lewandowski, D. (2008). The Retrieval Effectiveness of Web Search Engines: Considering Results Descriptions. Journal of Documentation, 64.

Lewandowski, D., Wahlig, H., and Meyer-Bautor, G. (2006). The Freshness of Web search engine databases. Journal of Information Science ,32(2): 133-150.

Mohamed, K. A.F. (2006). The impact of metadata in web resources discovering . Online Information Review, 30(2), 155 – 167. Retrieved June, 20, 2009, from Emerald database.

Namjoshi, N. (2004). Web Information Retrieval Using Web Document Structures [Abstract]. Master’s thesis. North Carolina State University . Retrieved Oct, 10, 2008, From http://wwwlib.umi.com/dissertations/fullcit/EP10011

Notess, G.R. (2003) Search Engine Statistics: Freshness Showdown . Retrieved Jan, 12, 2010, From http://www.searchengineshowdown.com/statistics/freshness.shtml

Qin, J and Wesley, K. (1998). Web Indexing with Meta Fields: a survey of web objects in Polymer chemistry. Information Technology and Libraries, 17(3), 149-156.

Safari, M. (2005). Search Engines and Resource Discovery on the web: Is Dublin Core an Impact Factor?. Webology, 2(2). Retrieved Apr, 20, 2010, From http://www.webology.ir/2005/v2n2/a13.html

Turner, T.P.; Brackbill, L. (1998). Rising to the Top: Evaluating the use of html meta tag to improve retrieval of world wide web documents through internet search engines. Library Resources and Technical Services, 24(4), 258-271.

Vaughan, L. (2003). New Measurement for Search Engine Evaluation Proposed and Tested . Information Processing and Management. Retrieved Oct, 11, 2010, From Elsevier Database.

Zhang, Z. ; Dimitroff, A. (2005b). The Impact of Metadata Implementation on Webpage Visibility in Search Engine Results (Part II) . Information Processing and Management, 41 (3), 691-715. Retrieved Feb, 18, 2005, From Elsevier Database.

Zhang, Z. ; Dimitroff, A. (2004). Internet Search engines response to metadata Dublin core implementation . Journal of Information Science , 30(4), 310-320. Retrieved Oct, 11, 2009, From Elsevier Database.

Zhang, Z. ; Dimitroff, A. (2005a). The Impact of Webpage Content Characteristics on

Webpage Visibility in Search Engine Results (Part I) . Information Processing and Management, 41 (3), 665-690. Retrieved Feb, 18, 2005, From Elsevier Database.

Zhang, Z. ; Jastram, I. (2005). A Study of the metadata creation behavior of different user groups on the Internet Information Processing and management. Retrieved Oct, 11, 2009, From Elsevier Database.